DOCUMENT RESUME

ED 080 313 SE 016 111

AUTHOR	Aikenhead, Glen S.
TITLE	The Interpretation of Student Performance on Evaluative Tests.
INSTITUTION	Saskatchewan Univ., Saskatoon. Dept. of Curriculum Studies.
PUB DATE	Mar 73
NOTE	19p.; Paper presented at the annual meeting of the National Association for Research in Science Teaching (46th, Detroit, Michigan, March 1973)
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	Academic Achievement; *Academic Performance; *Course Evaluation; Educational Research; *Nonparametric Statistics; *Physics; Science Education; *Secondary School Science
IDENTIFIERS	Project Physics; Research Reports

ABSTRACT

Reported is a study on the use of quantitative data in evaluating a science course for the purpose of introducing an alternative form of information presentation capable of supplying qualitative feedback valuable to students, teachers, and curriculum developers. Fifty-five teachers, randomly selected during the 1967-68 Project Physics (PP) experimental period, were classified as the PP trial, non-PP control, and PP experienced groups. Results of the pretest and posttest of the Test on Understanding Science and the Science Process Inventory taken by 1,188 randomly selected students, 921 in PP course and 267 in non-PP course, were used in this study. Quantitative t-test findings were considered as usable only in speculation of the nature of group differences. The McNemar chi square item analysis was introduced to identify areas of knowledge and specific ideas for which students showed a significant increase or decrease in understanding. The PP course was found superior to non-PP courses in four aspects: science tactics, value of science, instruction function of science, and science-society interaction. The increased specificity of qualitative analysis led to high-quality feedback. (CC)

Information and Research Report

Department of Curriculum Studies

College of Education
University of Saskatchewan
Saskatoon, Canada

THE INTERPRETATION OF STUDENT PERFORMANCE ON EVALUATIVE TESTS

Dr. Glen S. Aikenhead
Assistant Professor
Science Education

Paper presented at the 46th annual meeting of the
National Association for Research in Science Teaching
DETROIT, MICHIGAN
March 26 - 29, 1973

INTRODUCTION

This paper reports on. (1) the limitations to the use of quantitative data in evaluating a science course, and (2) a complementary or alternative form of information which yields valuable feedback to curriculum developers, teachers, and students. This feedback emerges from asking such qualitative questions as: What ideas have students learned? What misunderstandings have they still retained? Harvard Project Physics is partially evaluated in order to illustrate the use and limitations of qualitative data.

A NEED FOR QUALITATIVE DATA

Traditionally, student achievement and course evaluation rest firmly upon quantitative data. For example, "Students of course E gained 5.48 points on test T while students of course C gained only 1.71 points. Therefore, course E is better than course C."

But what does it mean to a curriculum developer or teacher for group E to score 3.77 points more than group C? What information does this give him? Unfortunately, it is an ambiguously summative datum which through statistical comparison leads to a probabilistic statement concerning course effectiveness. Most often such a statistical statement is misused as an interpretive and judgemental statement.[2] The education community's reliance on purely quantitative data tends to propagate this ambiguity and misunderstanding.

Quantitative data elegantly manipulated with sophisticated statistics do have their place. Welch and Walberg's multivariate analysis of variance study[3] is a paradigm in this regard. Their multivariate and univariate F-tests lead to useful comparisons between an experimental and control group. However, because these statistical computations rely on total test scores, the subsequent comparisons might very well be cursory. When identifying student achievement for the purpose of course evaluation, perhaps quantitative data should only be used as <u>initial</u> pieces of information.

One of the greatest utilities a test can have is supplying qualitative feedback to a researcher, teacher, and student.[4] Thus, one should consider asking qualitative questions such as: What ideas have students learned? What misunderstandings have they still retained? With the corresponding answers, the relative merit of one learning experience over another may be considered in less ambiguous terms.

Cooley and Klopfer[5] first demonstrated the assets of qualitative data when evluating Klopfer's History of Science Cases. Klopfer and McCann's[4] evaluation of a new junior high school science course and Klopfer's[6] evaluation of astronomy materials lent further credence to the use of qualitative data. Recently, MacKay[7] included qualitative data in a sur ey report carried out in Australia. The clarity of qualitative data allowed the researchers not only to conclude that one group of students "had learned more" than another group, but also allowedthem to describe in detail the ideas which the students had generally learned.

The dearth of qualitative data in educational research points to the pressing need for its inclusion in research studies.[8] Its utility is demonstrated here in an evaluation of Harvard Project Physics.*

---

*The first commercial edition was published as Project Physics, September, 1970; Holt, Rinehart & Winston.

AN EVALUATION OF HARVARD PROJECT PHYSICS*

A full description of the sample, experimental design, and instruments in the study may be found in another paper which is being presented to the 46th annual meeting of NARST.[1] Following is a summary of essential information. Fifty-five teachers were randomly selected from a total population of physics teachers in the United States and Canada. These teachers were then randomly assigned to teach Harvard Project Physics, HPP, (after having participated in a summer institute) or non-HPP (the physics courses they would have ordinarily taught.) An additional group of nineteen teachers, experienced at teaching HPP, volunteered to participate in the evaluation project. They taught in various regions of the United States. A random sample of students of all teachers wrote the Test on Understanding Science (TOUS)[9] or the Science Process Inventory(SPI)[10] on a pretest and posttest basis. These instruments do not concern themselves with what is normally called "subject matter." Instead

*The evaluation is only a partial one. The student learning evaluated in this study is limited by the two instruments which are used. To this extent, the evaluation of Harvard Project Physics is limited to knowledge about science and scientists.

they tend to measure one's knowledge about the scientific enterprise: the aims of science, its epistemology, its tactics, its values, its institutional functions, its interactions with society, and its human needs. There were 921 HPP and 267 non-HPP students. (The experimental group was larger than the control group in order to amass enough data for more extensive investigations with the HPP students.)

Results

The students' responses to the TOUS and SPI were analyzed both quantitatively and qualitatively.[11] According to statistical calculations, *t* tests for matched pairs, the TOUS and SPI scores gained significantly for both groups of students. These results are shown in Table I. The HPP students' mean gain score (7.95 points, TOUS and SPI combined) proved to be significantly greater than that of the non-HPP students (3.39 points, TOUS and SPI combined). However, with this quantitative information one can only speculate on the nature of these differences between the two groups. That is, the statistical comparison of quantitative data appear to be grossly uninformative.

TABLE I

TESTS FOR SIGNIFICANCE BETWEEN PRETESTS AND POSTTESTS

		Mean	T test*	Probability	SD**
HPP	TOUS (N=445)				
	pretest	34.41	12.51	p < .001	6.857
	posttest	37.54			7.059
	SPI (N=476)				
	pretest	107.52	10.65	p < .001	8.233
	posttest	112.34			8.245
non-HPP	TOUS (N=126)				
	pretest	35.25	2.83	.01<p<.001	6.434
	posttest	36.42			6.570
	SPI (N=141)				
	pretest	107.08	4.34	p < .001	7.789
	posttest	109.30			9.481

*Test for significance of the difference between two means for non-independent samples.

**SD means standard deviation.

A McNemar chi square item analysis[1] identified areas of knowledge and specific ideas for which students showed a significant increase or decrease in understanding. Table II summarizes the achievement of the HPP versus the non-HPP students. A significant increase or decrease in correct response between the pretest and posttest is indicated by a "+" or "-" respectively. The test items in Table II are categorized by their content. The arbitrary classification scheme is purely a heuristic one.

In four content areas:

(1) the tactics of science
(2) the values of science
(3) the institutional functions within science
(4) the interaction of science with society

the HPP achievement greatly exceeded that of the control group. Within these four categories the HPP students significantly improved on forty-three items compared with the control group's five.

Most notably, learning about tactics or methods of science appeared to be HPP's <u>tour de force</u>. All gains were accomplished by the HPP students alone. For example, a significant number of HPP students learned not to think in terms of "the scientific method," while

TABLE II

SUMMARY OF HPP VERSUS NON-HPP PERFORMANCE ON THE TOUS AND SPI (ITEMS CATEGORIZED WITH REGARDS TO CONTENT)[a]

Topic	Item[b]	HPP	Non-HPP
Aims of Science	S3C	-[c]	
	S20	+	+
	S86	+	+
	T12	+	
	T16	+	
	T23	+	
	S94	+	
	S15	+	
	S56		-
Epistemology of Science			
Definitions	S100	-	
	S13	+	+
	S40	+	
	S50	+	
	S62	+	
	S89	+	
	S58		+
Assumptions	S22	+	+
	S70	+	+
	S27	+	
	S17	+	
	S77	+	
	S114	+	+
	S119	+	+
	S12	+	
	S126	+	
	S109	+	
	S127	+	
	S73		-
	S104		+
	S44		+

TABLE II
(continued)

Topic	Item	HPP	Non-HPP
General aspects	T15	+	+
	S19	+	+
	S106	+	
	T10	+	
	T17	+	
	S125		-
	S63		+
Laws	S14	+	
	S49	+	
	S23	+	
Theories	T56	+	+
	S18	+	
	T57	+	
	S95	+	
	T45	+	
	S53	+	
	S110	-	
Models	S31	+	
	S33	+	
	T30	+	
	S59	+	
	S36	+	+
Observations	S5	+	+
	S113	+	
	S98	-	
	T51	-	
	S47	+	

TABLE II
(continued)

Topic	Item	HPP	Non-HPP
Tactics of Science			
"The scientific method"	S37	+	
	S82	+	
	S107	[illegible]	
	S121		
	S130	+	
	S135	+	
	T54	+	
Scientific inquiry	S115	+	
	S133	-	
	S112	+	
	S79	+	
	S48	+	
	T6	+	
	T13	+	
	S38	+	
	T55	+	
	S43	+	
	S26	+	
	T58	+	
	S88	+	
	S93	+	
	T26	+	
	T49	+	
Values of Science	S3	+	+
	S90	+	+
	S8	+	+
	S101	+	+
	S9	+	
	S55	+	
	S102	+	
	S97	+	
	S117	+	
	S6	+	
	S60	+	

TABLE II
(continued)

Topic	Item	HPP	Non-HPP
Instit[illegible]nal Functions			
[illegible]	T25	+	+
	S21	+	
	T59	+	
Other functions	T27	+	
	T2	+	
	T52	+	
	T28	+	
Interaction of Science With Society	T60	+	
	T29	+	
	T4	+	
Human Needs	T47	+	
	T18	+	
	T11	+	+
	T42	+	+
	S10	+	
	T9	+	
	T32	+	
	T31		+

[a] Included in this summary are only those TOUS and SPI items which showed a statistically significant McNemar chi square value for either the HPP or non-HPP groups.

[b] The number is preceeded by a "T" if the item comes from the TOUS, an "S" if from the SPI.

[c] A statistically significant McNemar chi value (.05 level of probability) is signified by the symbols "+" for an increase in correct response and "-" for a decrease in correct response.

non-HPP students apparently remained in their original state of misunderstanding. That is, only HPP students gained in their understanding that scientists are free to use any appropriate tactic in their research, "no holds barred" as Bridgman has said. Other topics subsumed under "tactics of science" included the roles of imagination, confirmation, and instrumentation in scientific methodology.

In the content area "values of science," items that showed significant improvement encompassed such topics as the simplicity value, skepticism, practical applications of scientific discoveries, and the values involved in reaching scientific conclusions. Knowledge of "institutional functions within science" generally increased for the HPP group alone. For example, these students became more aware of scientific journals, scientific societies, and the international character of science. The fourth content area to experience significant HPP gains was described as "the interaction of science with society." The items dealt with the limitation of scientific thought applied in non-scientific areas, the influence of government upon scientific progress, and the effect of public support upon science.

The HPP group significantly gained on twice as many items as the non-HPP group in the remaining three areas:

(1) the aims of science
(2) the human needs within science
(3) the epistemology of science -- including definitions, assumptions, general aspects of scientific knowledge, laws, theories, models, and observations.

Even though the HPP and non-HPP courses appear more similar in these content areas, the HPP course did seem to encourage greater achievement.

Using the data in Table II and referring to the pertinent test items, one can compare the HPP and non-HPP groups in terms of the specific ideas students tended to learn during the year. An illustrative complex comparison is made for the section "aims of science." According to SPI items 20 and 86, most physics students regardless of the course of study tended to learn that science strives to establish casual relationships which are not necessarily more complex than previous relationships. However, <u>only</u> the HPP students showed significant improvement in: (1) learning the aim of scientific explanation (TOUS items 12 & 16), (2) distinguishing between science and technology (TOUS item 23),

(3) recognizing the importance of prediction as a goal in science (SPI item 94), and (4) realizing that scientists do not wish to make prejudiced observations (SPI item 15). On the other hand, some HPP students seemed to be misled into thinking that the purpose of experimentation is to prove the laws of nature (SPI item 30), while the non-HPP students acquired the misconception that scientific investigations may be exemplified by the simple activity of collecting rocks (SPI item 56).

Summary

Compared with other physics courses, the impact of HPP on students learning about science and scientists appeared to be substantial. The HPP students achieved significantly greater gain scores than did the non-HPP students. While HPP did not clarify misconceptions and misunderstandings in the minds of all students, it did tend to improve (much more than did other physics courses) the student's knowledge about science and scientists. In four areas specifically (the tactics of science, its values, its institutional functions, and the interaction of science with society) the HPP achievement <u>greatly</u> exceeded that of the control group.

IMPLICATIONS FOR FURTHER RESEARCH

Quantitative data do play a logical role in determining statistical conclusions for experimental studies. However, this role might best be limited to a minor one when identifying student achievement for the purpose of course evaluation. Qualitative data appear to be far more informative to curriculum developers and teachers than total score gains and subscore gains alone. In this study, the evaluation of HPP demonstrates the increased specificity in the results derived from qualitative information. Not only can one say the HPP students achieved more than the non-HPP students, but one can better understand in what ways (by what knowledge) the groups differed. In some measure, the developers and teachers of HPP may now understand how their course differs or concurs with other physics courses. They can realistically anticipate what learning will likely take place. If this content is thought to be incomplete, the appropriate changes may be made to the course.

In addition to comparing experimental and control groups, one can also use the qualitative data to compare the ideas generally learned by students with the expressed objectives of the course. Such a study has been completed for HPP.[11]

REFERENCES

1. Aikenhead, Glen S. "A New Methodology for Test Construction in Course Evaluation." A paper persented at the 46th annual meeting of the National Association for Research in Science Teaching, Detroit, March 26-29, 1973.

2. Heath, Robert W. "Pitfalls in the Evaluation of New Curricula." Science Education, 46 (April, 1962), 216.

3. Welch, Wayne W., and Walberg, Herbert J. "A National Experiment in Curriculum Evaluation." American Educational Research Journal,9 (Summer, 1972), 373-383.

4. Klopfer, Leo, and McCann, Donald C. "Evaluation in Unified Science." Science Education, 53 (March, 1969), 155-164.

5. Cooley, William W., and Klopfer, Leo. "The Evaluation of Specific Educational Innovations." Journal of Research in Science Teaching, 1 March, 1963), 73-80.

6. Klopfer, Leo. "Effectiveness and Effects of ESSP Astronomy Materials." Journal of Research in Science Teaching, 6 (January, 1969), 64-75.

7. MacKay, Lindsay D. "Development of Understanding About the Nature of Science." Journal of Research in Science Teaching, 8 (January, 1971), 57-66.

8. Aikenhead, Glen S. "The Measurement of High School Students' Knowledge About Science and Scientists." Paper presented at the 44th annual meeting of the National Association for Research in Science Teaching, Silver Spring, Maryland, March 25, 1971.

9. Cooley, William, and Klopfer, Leo. Test on Understanding Science. Princeton, New Jersey: Educational Testing Service, 1961.

10. Welch, Wayne W. Welch Science Process Inventory, form D. Cambridge, Mass.: Harvard Project Physics, 1966.

11. Aikenhead, Glen S. "The Measurement of Knowledge About Science and Scientists: An Investigation into the Development of Instruments for Formative Evaluation." Unpublished Doctoral Thesis, Harvard University, 1972, pp. 95-194.